

## Review: Histograms & Describing

Using your calculator, create a histogram of the following data, then describe the distribution.

20	20	21	23	24	22	25
25	25	25	26	28	30	34
38	40	46				

## 2.2: More specifics:

### Center & Spread

#### CENTER: 2 choices: Median or Mean

##### 1) MEDIAN:

- put numbers in order
- find the middle number
- if there are two middle numbers, average the two

Example:

3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20

~~3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20~~  
~~3, 3, 4, 5, 6, 6, 7, 8, 10, 11, 12, 15, 18, 20~~

Median = 7.5

#### QUARTILES:

- the medians of the lower and upper halves of the data
- first and third quartiles
- Q1 and Q3

Same example:   
3, 3, 4, 5, 6, 6, 7, 8, 10, 11, 12, 15, 18, 20  
Q1 = 5      Q2 = 7.5      Q3 = 12  
25%      25%      25%      25%

SPREAD: The spread that goes with the Median is the IQR and full range

RANGE: (min, max)

IQR: Inter Quartile Range

- the difference between Q3 and Q1

IQR = Q3 - Q1 = # units

- the middle 50% of the data

Same example: 3, 3, 4, 5, 6, 6, 7, 8, 10, 11, 12, 15, 18, 20

Q1 = 5

IQR = 7 units

Q3 = 12

Range = (3, 20) units

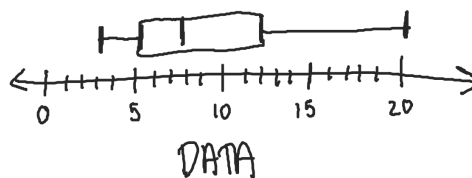
#### THE 5# SUMMARY: Center & Spread together

Min, Q1, Median, Q3, Max

#### Visually: BOXPLOT

Use the data from before: 3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20

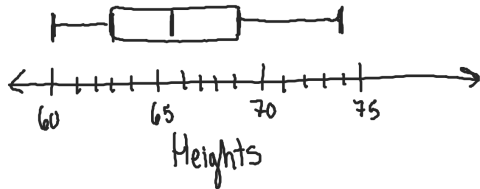
Min = 3      Q1 = 5      Median = 7.5      Q3 = 12      Max = 20



Example: Create a boxplot of the following:

Heights of a sample of HS students:

Min = 60" Q1 = 63" Med = 65.5" Q3 = 69" Max = 74"



5 # Summary and Boxplots on the calculator:

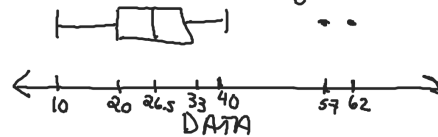
Find the 5# Summary & create a boxplot of the following data:

12	18	25	20	21	17	30	29	32	35
10	40	38	34	33	31	30	25	24	23
19	22	13	16	26	29	27	35	62	57

See page 71--72 in the book

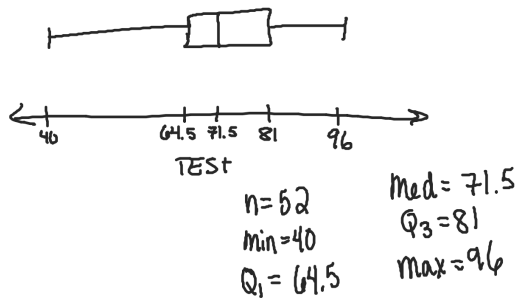
STAT --> CALC--> 1-var stats --> Find the list you want

$\bar{x}$  = average, mean

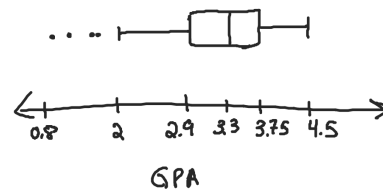


### PRACTICE:

1) Create a boxplot and find the 5# summary of the list TEST



2) Create a boxplot and find the 5 # Summary of the list GPA



### Finding outliers- the 1.5 x IQR test

In order to formally to determine if there are outliers:

- find  $1.5 \times (IQR)$
- Take this number and:
  - subtract from Q1  $Q1 - (1.5 \times IQR) = A$
  - add to Q3  $Q3 + (1.5 \times IQR) = B$
- This is considered the range of acceptable data (A, B)
- ANYTHING OUTSIDE THIS RANGE IS CONSIDERED AN OUTLIER

Example: Using the data from before, test for outliers

12	18	25	20	21	17	30	29	32	35
10	40	38	34	33	31	30	25	24	23
19	22	13	16	26	29	27	35	62	57

5# Summary:

min = 10

Q1 = 20

Med = 26.5

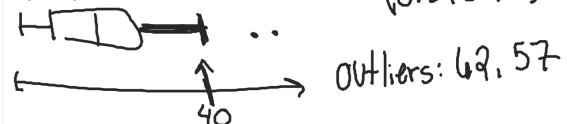
Q3 = 33

Max = 62

$$A = 20 - (1.5 \times 13) = 0.5$$

$$B = 33 + (1.5 \times 13) = 52.5$$

(0.5, 52.5)



Example: Create a boxplot (testing for outliers) of the following data. Put the data in L2 in your calculator.

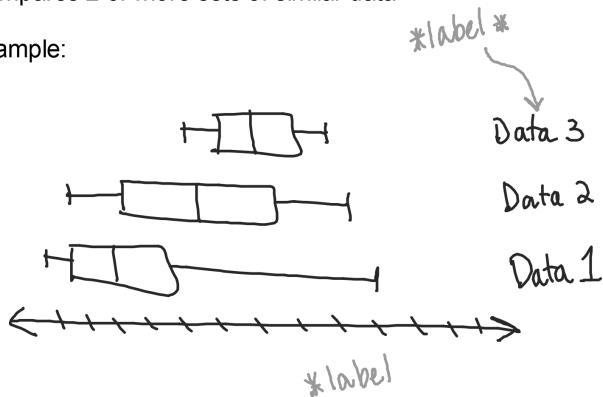
10	45	60	62	55	50	49	51	52	55
69	67	63	64	66	65	70	72	73	48
46	47	53	54	57	58	59	60	70	8
60	56	47							

## 2.2 practice worksheet A

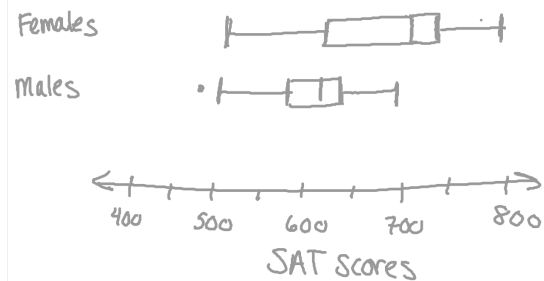
### PARALLEL BOXPLOTS:

\* Compares 2 or more sets of similar data

\* Example:



Example: Create parallel boxplots of SATMF and SATMM



### Review:

Measure of Center: Median

Measure of Spread: IQR & Range

On calculator: 1-Var Stats

Picture: Boxplot, showing outliers

The other center...

### CENTER: the Mean

\* Arithmetic average

\* Add all observations together and divide by the number of observations.

\* symbol:  $\bar{X}$

Med  
M > median

\* Read as "x bar"

\* **Must be written with the bar!!!!**

\* on calculator: 1-var stats....  $\bar{X}$  (1<sup>st</sup> number)

**Spread:** When we have chosen Mean as the center...

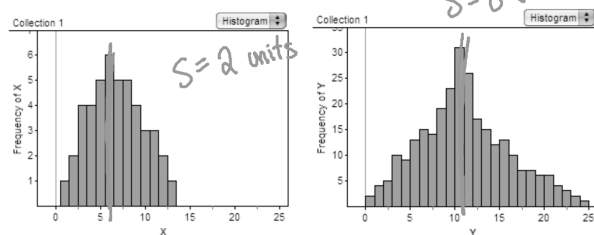
\* Standard Deviation =  $\bar{x} = 65$

A number that describes how spread out the data are.

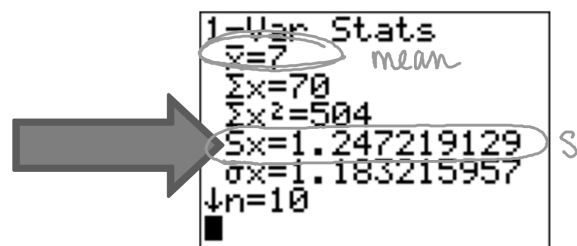
\* Average difference of each point from the mean.

\* The higher the number the more spread out the data

\* symbol:  $S$



Just find it on the calculator:



Look at the two sets of data and decide which one has the larger std. deviation:

1)  $\{4, 9, 12, 17, 20\}$  or  $\{4, 6, 7, 8, 10\}$

2)  $\{100, 140, 150, 160, 200\}$  or  $\{10, 50, 60, 70, 110\}$  *same*

Examples: Which one has the largest/smallest std. dev?

1) 4, 5, 6, 6, 7, 7, 7, 8, 8, 9, 10



2) 4, 4, 4, 5, 5, 6, 7, 8, 9, 9, 10, 10, 10

$S = 2.449$



3) 1, 3, 5, 7, 7, 9, 9, 9, 11, 11, 13, 15, 17

$S = 4.546$



4) 5, 5, 5, 5, 5

$S = 0$

### Properties of the standard deviation (s):

\*  $s = 0$  when... All points are the SAME

\* Unless all points are the same,  $s$  is ALWAYS ... POSITIVE

\* It is affected by outliers

$s \uparrow$



Summary:

CENTER:

MEDIAN

MEAN

SPREAD:

RANGE  
IQR

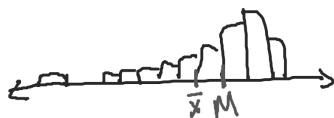
RANGE  
STD. DEVIATION

## Mean vs Median (in distributions)

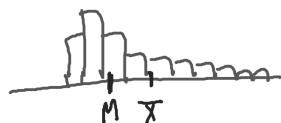
\* Symmetric:  
mean  $\approx$  median



\* Left skewed:  
mean < median

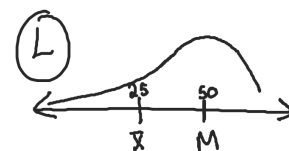


\* Right skewed:  
mean > median



## Examples:

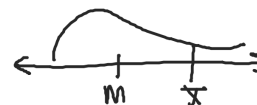
1) Mean = 25  
Median = 50



2) Mean = 300  
Median = 305

Symm.

3) Mean = 240  
Median = 200



\* The mean is affected by outliers/skew

\* The median is resistant to outliers/skew



So when do we use each?

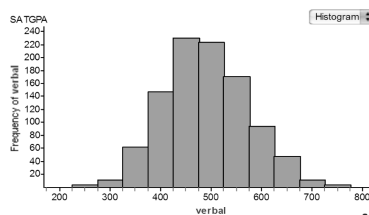
\*\* Symmetric data => Mean and Std. Deviation

\*\* Skewed data (or outliers) => Median and IQR



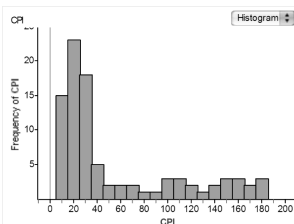
## Examples:

Let's describe these distributions again.



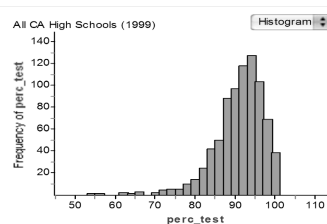
mean = 490  
std dev = 90  
min = 225  
Q1 = 410  
Med = 495  
Q3 = 600  
max = 775

- symmetric, unimodal
- Center @ mean of 490 pts.
- std. dev. of 90 pts.
- range of (225, 775) pts.



mean = 63  
std dev = 20  
min = 10  
Q1 = 30  
Med = 41  
Q3 = 100  
max = 185

- Right skewed, unimodal
- Center @ Median of 41 units.
- IQR of 70 units
- Range of (10, 185) units.



mean = 85  
std dev = 8  
min = 55  
Q1 = 83  
Med = 91  
Q3 = 96  
max = 102

- left skewed, unimodal
- Center @ Med. of 91 units
- IQR of 13 units
- range of (55, 102) units
- Gap @ approx 58 and 68

#2 | mean: increase  
median: increase slightly, not much  
std.dev: increase

#3 | a)  $\bar{x} = \$18,000$      $M = \$15,500$

b) outlier

c) Median + IQR

d)  $S = \$11,53.47$

## Center and Spread worksheet

### Comparing Distributions:

Example: Compare & Describe the heights of males & females

\* When comparing 2 distributions, you MUST still state **shape**, **center and spread** of both distributions

\* You must also **compare** these 3 things in each distribution

\* Use comparison words: higher, lower, similar, same, wider, etc.

\* Cannot compare mean to median **any outliers/skew**  
**std.dev to IQR**

Example: Block 2 & 3 took a test (out of 100 points).

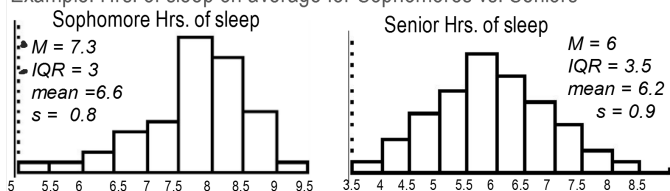
\* The shape of Block 2 is Sym which is similar to (different than) the shape of Block 3 which is left skew. Both are unimodal.

\* The center of Block 2 is the mean/median of 75% which is higher/lower than the mean/median of Block 3 which is 80%.

\* The IQR std. dev. of Block 2 is 5% which is wider/smaller than the IQR std. dev. of Block 3 which is 10%. The range of Block 2 is (\_\_, \_\_) which is wider/smaller than the range of Block 3 which is (\_\_, \_\_).

\* Other??? outliers? gaps?

Example: Hrs. of sleep on average for Sophomores vs. Seniors



- Soph. are left skewed which is diff. than seniors who are symmetric. Both are unimodal.
- The soph Median is 7.3 hrs, which is higher than Senior Med. of 6 hrs.
- The soph IQR is 3 hrs, which is narrower than Senior IQR of 3.5 hrs.
- The soph range is (5, 9.5) hrs, which is narrower than Seniors range of (3.5, 8.5) hrs.

Example: Compare and Describe the SATMM and SATMF lists.

Create the histograms (on your calculator only) & find the 1-Var Stats before you start.

**SATMF**  
 left skew  
 unimodal  
 $M = 700$   
 $IQR = 120$   
 $r: (500, 800)$

**SATMM**  
 right skew  
 unimodal  
 $M = 610$   
 $IQR = 55$   
 $r: (480, 700)$

- Females are left skewed, which is different than males which are right skew. Both are unimodal.
- Females Med. is 700<sup>pts</sup> which is higher than males Med. of 610 pts.
- Females IQR is 120 pts, wider than males IQR of 55 pts.
- Females range is (500, 800) pts, wider than males range of (480, 700) pts.
- Males have an outlier @ 480 pts.

**Compare & Describe the BLK1 and BLK2 lists. These are quiz scores out of 20 points.**

Steps:

1- Create the histograms (on the calculator only) to determine shape

2- Find the 1-Var Stats and jot them down somewhere.

3- start describing (shape, center, spread) *comparison words*

4- Check boxplots on the calculator to see if there are any outliers- be sure to mention these in your description!

BLK1 *skewed left, med = 15, IQR = 5, range (4, 20)*

BLK2 *roughly symm or skew right, med = 13, IQR = 3.5, range (8, 18)*

\* Block 1 is skewed left, whereas Block 2 is symmetric (or slight right skew). Both are unimodal.

\* Block 1 has a median of 15 points, which is higher than the Block 2 median of 13 points.

\* Block 1 has an IQR of 5 points, which is higher than Block 2 of 3.5 points.

\* Block 1 has a range of (4, 20) points which is wider than Block 2 of (8, 18) points.

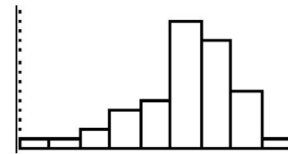
\* Block 1 has an outlier at 4 points.

**Compare & Describe the NONSM and SMOKE lists. These are cholesterol levels of Non-Smokers and Smokers. Cholesterol level is measured in points.**

**Do this on separate paper so you can turn it in!**

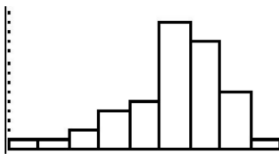
**Outliers: How do they affect summary statistics?**

Original data:



If I add an upper outlier, how will it affect the summary stats?

Mean  
Median  
range  
IQR  
Std. deviation



If I add a lower outlier, how will it affect the summary stats?

Mean  
Median  
range  
IQR  
Std. deviation

\* Outliers do not really affect the Median, IQR, or Quartiles

\* Outliers STRONGLY affect the mean, standard deviation, and range

**Example:** Workers at a local business make on average \$30,000 per year. A new employee is hired to be the company President, and he will make \$70,000 per year.

What will happen to the following summary statistics if the President's salary is added to the group?

Mean  
Median  
Range  
IQR  
Std. Deviation  
Q1

**Example:** Workers at a local business make on average \$30,000 per year. A new employee is hired to be part time only and he will make \$10,000 per year.

What will happen to the following summary statistics if the new employee's salary is added to the group?

Mean  
Median  
Range  
IQR  
Std. Deviation  
Q1

*Try Classwork 2.2B*